Reprinted without change of pagination from the
Proceedings of the Royal Society, A, *volume* 299, *pp.* 291–306, 1967

On the theory of Hertzian fracture

By F. C. Frank, F.R.S. and B. R. Lawn
H. H. Wills Physics Laboratory, University of Bristol

(*Received* 30 *August* 1966—*Read* 20 *April* 1967)

[Plate 2]

The fracture of a brittle solid under a spherical indenter is the best studied case of fracture in a strongly inhomogeneous, well defined, stress field. Two principal topics are discussed, the path of a crack in a field of non-uniformly directed stress, and the stability of cracks of various length when the prior stress on the crack path is non-uniform. For the first, it is shown that the crack growth should, *to a first approximation*, be orthogonal to the most tensile principal stress, and thus correspond, in a torsion-free stress field, to a surface delineated by the trajectories of the other two principal stresses: while, to a second approximation, the crack should deviate from this path by having a larger radius of curvature at every bend, thus exhibiting a pseudo-inertia even in slow growth. This is in accordance with the known experimental facts about the Hertzian crack, particularly the fact that the crack at the surface forms systematically outside the edge of the circle of contact, at which the maximum tensile stress occurs. On the second question, it is found that there are four crack lengths, c_0, c_1, c_2, c_3, corresponding to stationary values of energy. c_0 and c_2 represent unstable equilibria, and diminish with increasing load; c_1 and c_3 represent stable equilibria and increase with increasing load. With small indenters, c_0 soon becomes less than the size of pre-present surface flaws, and an unobserved shallow ring crack of depth c_1 is produced: the critical condition for observed fracture is then the merging of c_1 with c_2, allowing unstable growth to the cone crack of depth c_3. This explains Auerbach's law, that the critical load for production of a cone crack is proportional to the radius, r, of the indenter sphere. With larger indenters, of several centimetres radius for a typical case, c_1 and c_2 merge and disappear before c_0 exceeds the size of pre-present flaws. The critical load for cone fracture then becomes nearly proportional to r^2, as observed. The previous calculations of Roesler (1956*a*, *b*) relate to the second stable crack dimension, c_3, though his energy scaling principle is also applicable to the critical condition at which c_1 and c_2 merge. The Hertzian fracture test, within the validity range of Auerbach's law, affords a means of measuring surface energy at the fracture surface independent of knowledge about the pre-present flaws.

1. Introduction

If a hard spherical indenter is pressed with an increasing normal force P on to the flat face of a brittle solid, a cone crack develops when P reaches a critical value P_c. If this experiment is repeated with a number of spheres of various radii r, the materials being the same, it is empirically found that P_c is proportional to r. Following the suggestion of Roesler (1956*a*) we refer to this fact, first found by Auerbach (1891), as 'Auerbach's law'. In conjunction with Hertz's (1881) calculation of the elastic stresses under this type of loading, it implies that the maximum tensile stress in the body, just prior to fracture, is proportional to $r^{-\frac{1}{3}}$ instead of a constant value, as one might have supposed. One suggested explanation of this size effect is based on Griffith's (1920) demonstration that in most brittle substances fracture begins from pre-existing flaws, with the consideration that a larger stressed area has a greater chance of including a particularly weak flaw. Roesler, however, refutes this explanation on the grounds that the scatter in results is just about the same with large and small indenters.

Hertz noted that the duration of an impact, for example, with a bouncing sphere, is sufficient to make the quasistatically evaluated stress a good approximation within the region in which it is at all large. This enables Roesler (1956*a*) to demonstrate the equivalence of a number of different empirical formulations of the critical conditions of Hertzian loading, either slowly applied, or by impact, for production of a cone crack. All can be brought into the form that just before fracture the elastically stored energy is proportional to the area of the circle of contact. Cone cracks are geometrically similar, scaling with the circle of contact, and Roesler suggests that the fundamental form of the law is proportionality between elastically stored energy and crack surface area.† The law would be explained, he suggests, if we suppose that the crack grows nearly reversibly, the small fraction of the elastically stored energy released by crack growth just about balancing the increase of surface energy throughout the process, finally amounting to a size-independent fraction η of the prior total stored energy. We shall propose another interpretation, which does not imply near reversibility in the main growth of the crack.

2. The stress field

We base our interpretation on the special nature of the inhomogeneous stress field of Hertzian loading (Hertz 1881, 1882). The radius a of the circle of contact between spherical indenter and flat specimen, and the maximum tensile stress σ_m in the specimen, are found from the Hertzian analysis:

$$a^3 = \tfrac{4}{3}kPr/E, \tag{2·1}$$

$$\sigma_m = (1-2\nu)P/2\pi a^2 = \tfrac{1}{2}(1-2\nu)p_0, \tag{2·2}$$

$$k = \tfrac{9}{16}[(1-\nu^2)+(1-\nu'^2)E/E'], \tag{2·3}$$

where P is the normal load on the indenter, r is the indenter radius, E and E' are the Young modulus of specimen and indenter respectively, and ν and ν' the corresponding values of the Poisson ratio. p_0 is the *mean* pressure acting between indenter and specimen and serves as a unit of stress. The factor k conveniently becomes unity, for $\nu = \frac{1}{3}$, if indenter and specimen are made of the same material. Figure 1, drawn for $\nu = \frac{1}{3}$ (we assume this value for all numerical examples presented in this paper) shows the value of the greatest tensile stress σ_1 (see below) as a function of position in a plane through the axis of symmetry. There is a drop-shaped region below the area of contact in which all three principal stresses are compressive. Outside this region, and remote from the free surface, SS, the greatest tensile stress rises to a modest value of about $0{\cdot}005p_0$, and falls off again with distance. Much larger tensile stresses, in excess of $0{\cdot}01p_0$, occur in a shallow region near the free surface outside the area of contact, and reach their maximum σ_m *at* the circle of contact (AA in figure 1).

† From this it follows that the Auerbach constant, P_c/r, for a static test, is replaceable for dynamic tests, when W_c is the kinetic energy of the impacter, according to

$$P_c/r = (\tfrac{5}{2})\,(9E^2W_c^3/100h^2r^4)^{\frac{1}{5}}$$

which is equation (12) of Roesler (1956*a*) rewritten in the notation defined in §2 below.

The directions of principal stresses are best displayed by the stress trajectories (cf. Morton & Close 1922).† We choose to label the three families of stress trajectories in such a way that in general $\sigma_1 > \sigma_2 > \sigma_3$. These respectively replace the p_2, p_3, p_1 of Morton & Close. σ_2 is a 'hoop-stress', its trajectories being circles about the symmetry axis. σ_1 and σ_2 trajectories form a pair of orthogonal families in any plane through the symmetry axis. A σ_3 trajectory is approximately a hyperbola meeting the surface orthogonally (the stress σ_3 there falling to zero if the point lies outside the circle of contact) and having an asymptote radial from a point somewhat above the

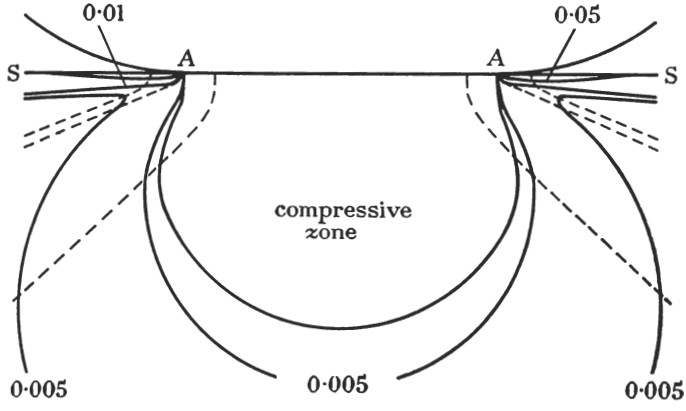

FIGURE 1. Contours of greatest principal stress, σ_1, in semi-infinite elastic medium (surface SS) in contact (diameter of contact AA) with spherical indenter. p_0 is taken as the unit of stress. Broken lines are σ_3 stress trajectories drawn from surface at distances $0{\cdot}8a$, a, $1{\cdot}2a$ from centre of circle of contact.

centre of the circle of contact. Three examples, starting from $0{\cdot}8a$, a, and $1{\cdot}2a$ at the surface, are shown in figure 1. σ_1 trajectories approximate at depth to circles about this centre, and have inflexions where they enter the region of relatively high stress near the free surface, where they turn rather rapidly towards parallelism with this surface. σ_1 is always the most tensile of the three principal stresses, and the greatest tensile stress values shown in figure 1 are therefore σ_1 values throughout. σ_2, which is tensile except near or within the compressive zone, is the intermediate stress except in a small region near the surface, just within the circle of contact (between $0{\cdot}97a$ and a), in which $\sigma_3 > \sigma_2$. σ_3 is everywhere compressive.

† Hertz calculated the stress trajectories approximately within a limited region, and his extrapolation outside this region was erroneous. They were re-computed by Fuchs (1913). Fuchs's diagram purports to show the stress trajectories in a sphere, rather than a plane-surfaced body, and is, in consequence, in error: as he himself remarks, within the validity of the approximations used, if the scale is chosen so that the circle of contact appears with finite radius, the radius of the sphere should appear infinite. As a result, his stress trajectories of the family which we denote as σ_1 actually meet the free surface at a finite angle, instead of extending asymptotically to this surface, and his σ_3 trajectories (our notation) do not meet the free surface quite orthogonally. These errors have become progressively exaggerated in the copying of Fuchs's diagram into Love (1927) and Haward (1949). Morton & Close show the correct general form of a σ_1 trajectory (p_2, in their notation) but one at least of their σ_3 trajectories (p_1) must have been rather carelessly sketched, not crossing the σ_1 trajectory orthogonally.

3. The crack path

The σ_3 stress trajectory leading from σ_m in figure 1 bears a remarkably close resemblance to a section of the fully developed cone crack. The chief empirical discrepancy is that the crack generally meets the free surface a little *outside* the periphery of the circle of contact, rather than at it. On the other hand, the angle α which this σ_3 trajectory asymptotically makes with the symmetry axis, $\sim 68°$, is in remarkably close agreement with that observed experimentally ($68{\cdot}5 \pm 1°$ according to Roesler 1956*b*). Thus to an excellent first approximation we may say that cracking proceeds orthogonally to the greatest tensile stress σ_1, thus following a surface delineated by the trajectories of the other two principal stresses, σ_2 and σ_3. Starting at the free surface from a point of maximum σ_1 (for example, at $\sigma_1 = \sigma_m$), the σ_2 trajectory carries the crack round in a circle about the axis, and the σ_3 trajectory carries it downwards, initially vertically and subsequently into an ever-widening cone. (This is the *principal* or *primary* crack: there are *secondary* cracks which we need not discuss till later.)

One may express surprise that the crack path, apart from its place of commencement, is in any simple way related to the state of stress in the body prior to cracking: for the further growth of a crack beyond any particular stage in its growth is assuredly determined by the extremely different stresses present at that stage. A little consideration, however, will show that a close relation between stress trajectory and ultimate crack path, as outlined above, must exist, although the correspondence will not be exact.

We note that ΔU, the reduction in mechanical energy (the sum of elastically stored energy and energy of the loading system) due to formation of the crack, is calculable as the work required to be done by tractions applied to the crack faces to bring them together again. Starting with the crack in elastic equilibrium at any stage in its growth, a workless constraint to prevent further growth being applied at the crack front, one applies tractions to its faces which increase linearly and proportionally until they reach values equal to the stresses along the crack path as they were prior to fracture. The *strains* are therefore also precisely as they were before fracture, and the crack is closed, with corresponding points on its two faces precisely juxtaposed again. The energy decrease caused by crack formation is therefore expressible in terms of an integral of prior stress multiplied by subsequent relative displacements of corresponding points, over the crack surface. These displacements are, in principle, calculable from the tractions on the crack surface and the elastic properties of the system. Thus, for a given system, ΔU at any stage of crack growth depends on the crack path up to that stage and the prior stresses on that path.

The location and orientation of the next increment of crack growth must, for growth under reversible conditions where the release of mechanical energy only just suffices for further propagation of the crack, be that which maximizes the quantity $\mathrm{d}(\Delta U)/|\mathrm{d}\mathbf{c}| - 2\gamma$, where $\mathrm{d}\mathbf{c}$ is an incremental area of crack and γ is the cost in free energy of making unit area of new surface. (Anisotropy in γ plays a significant role in determining the path of crystal cleavage, but we are now considering such

materials as glass, for which we shall regard γ as independent of position and orientation.) It is likely that the above condition holds so long as the speed of crack propagation is not a large fraction of the speed of transverse sound waves.

The quantity $\mathrm{d}(\Delta U)/|\mathbf{dc}|$ may be calculated in two ways: from the prior stresses, as the difference between the amounts of work required to close the whole crack with and without the increment $\mathbf{dc}$: or from the actual stresses at the location of $\mathbf{dc}$, as the amount of work required to close the incremental crack alone. The calculation of this term is indeed the basis of Irwin's 'fracture mechanics' approach (Irwin 1958) to the problem of crack extension, about which more will be said in § 4. Griffith (1924), dealing with the problem of an obliquely stressed crack in plane strain or plane stress, postulates that the crack extends from that place adjacent to the crack tip (regarded as having an elliptical cross-section) where the actual (as distinct from prior) tensile stress is greatest. The Griffith treatment carries the implication, from the way he relates this problem to his former one (1920) of a perpendicularly loaded crack, that the crack extends orthogonally to this maximum tensile stress. Using the fact that the energy reduction due to a crack, of given size and of arbitrary orientation in a stressed body, is greatest when the crack plane is perpendicular to the greatest principal tensile stress and is insensitive to the values of the other two principal stresses, one concludes that the location and orientation of the crack increment as postulated by Griffith is indeed that which maximizes $\mathrm{d}(\Delta U)/|\mathbf{dc}| - 2\gamma$. It is to be noted that the path maximizing this quantity at each increment is not necessarily the path of maximum overall energy release; nor is it necessarily the path of maximum prior stress, nor of maximum normal component of tensile stress: cf. figure 1, which shows that the actual crack path, which closely approximates the σ_3 trajectory beginning from the circle of contact AA, is very far from satisfying either of the latter conditions.

From the foregoing argument we conclude that while the crack path can be regarded as controlled by the prior stresses, it is influenced by these stresses over the whole of the previous path and not only by the prior stresses at its growing edge. We must therefore deny the hypothesis that it follows stress trajectories *exactly*. Quite clearly, if a crack growing perpendicularly to a strong uniaxial tension passes into a region of zero stress, it will continue for some distance on the same plane, and the result will be little different if the second region is weakly stressed in any direction, so that arbitrarily large deviations between the crack and the surface defined by the lesser principal stresses are possible. In a more regularly varying stress field, so long as the trajectories of the two lesser principal stresses define a plane, a crack on that plane will continue on it. When the surface they define bends the crack will tend to continue on its former plane, as if it possessed inertia, though we do not invoke a dynamic cause. It thus enters a region in which it is obliquely stressed, and thereby suffers deflexion, in the direction readily inferred from Griffith's case, namely towards perpendicularity to the new orientation of greatest principal tensile stress.†

† There are some complex possibilities in the general case. One is that in a stress field with torsion there is no unique surface defined by following the trajectories of the two lesser principal stresses from one point, so that the form of the crack on its growth surface must be

Footnote continued on next page

Thus we may say, at least qualitatively, that the crack follows the surface defined by the σ_2 and σ_3 stress trajectories in much the same way as a massive suspended particle follows the flowlines of an airstream, swinging wide on the bends: the crack will curve in the same direction as the stress trajectory surface, but with an exaggerated radius of curvature.

We may now re-examine the empirical evidence with regard to Hertzian cracks to see whether these anticipations are confirmed. Observers agree that the diameter of the Hertzian cracks at the specimen surface exceeds that of the circle of contact (by an average of 19 % according to Auerbach (1891) or 12 % according to Tillet (1956)). Roesler, after rejecting the explanation of Auerbach's law by a statistical distribution of flaws, accepts this explanation for the excessive diameter of the surface crack. It would alternatively be explained if the crack began at a place at or close to the edge of the circle of contact, where the stress is greatest, and then spread around, not exactly on the σ_2 circle, but with a somewhat greater radius of curvature. We give reasons below for supposing that the crack first runs around to form a shallow ring and only subsequently extends in depth. This explains the excessive diameter of the ring crack at the surface. The pseudo-inertia of crack growth has a compensatory effect when the crack deepens: for, if it strictly followed the σ_3 trajectories from the actual surface ring we should expect it to finish with a larger semiangle than the 68° which corresponds to σ_3 trajectories from the circle of contact. Because of the pseudo-inertia, it should dive deeper, so that the two deviations tend to compensate each other in their effect on the cone angle. According to these views, the surface ring crack should be somewhat eccentric relative to the circle of contact, and the cone axis should not truly coincide with the surface normal.

4. Critical conditions for cone fracture

Figure 2 shows the distribution of stress σ_1 normal to the downward crack path as a function of the distance c along it, assuming that the path follows precisely along the σ_3 trajectory commencing at the circle of contact. As indicated in § 3 this assumed path is inexact. In particular, the initial stress (at $c = 0$) would be only about $0{\cdot}8\,\sigma_m$ starting at a point $1{\cdot}12a$ from the axis. However, as we have pointed out, we believe that in general the ring crack initiates close to the circle of contact. This most highly stressed portion of the ensuing eccentric ring crack should act as the leader, determining the conditions for the ultimate penetration of the whole crack. In any case the main qualitative features of the stress distribution along the

known before the surface on which its further extension should occur can be predicted. Another is that gross instability of the crack path would set in after a change in dominance of the principal stresses, one of the hitherto lesser principal stresses becoming the greatest. Such a state of stress is produced, for example, when a cylindrical sleeve is shrunk tight on to a mandrel held in longitudinal tension. A similar state of stress is apparently produced spontaneously by over-rapid cure of a fairly thick cylinder of polyester resin. The transition from simplicity to ordered complexity which occurs in such specimens where a crack from the surface enters the core region deserves aesthetic no less than scientific contemplation, and receives both from Professor L. O. Nicolaysen of the Bernard Price Institute for Geophysical Research, Johannesburg, by whom they were shown to one of us. Fortunately neither complication is present in the Hertzian stress field.

path assumed in figure 2 are similar for all alternative paths within the range of uncertainty. These are that the stress rapidly falls approximately linearly within a small distance $\delta \ll a$, after which it gradually diminishes, approximately as $(c+\frac{1}{2}a)^{-1}$. This gradual fall-off is so slight between $c = \delta$ and $c \simeq 0{\cdot}3a$ that the stress may be regarded as constant over this range. From figure 2 we may write the total stress as

$$\left.\begin{aligned} &\sigma_1 = \sigma' + \sigma'', \\ \text{where} \quad &\sigma' = 0{\cdot}06\,\sigma_m \\ \text{and} \quad &\sigma'' = (\sigma_m - \sigma')\,[1-(c/\delta)] \quad (c < \delta), \\ &\sigma'' = 0 \qquad\qquad\qquad\quad (c \geqslant \delta), \end{aligned}\right\} \tag{4·1}$$

and δ is related proportionally to a according to

$$\delta = 0{\cdot}02a. \tag{4·2}$$

σ_m is defined in (2·2). The parameters a and δ are convenient scaling lengths for the stress field: we should note that they are not constant, but increase with load according to (2·1).

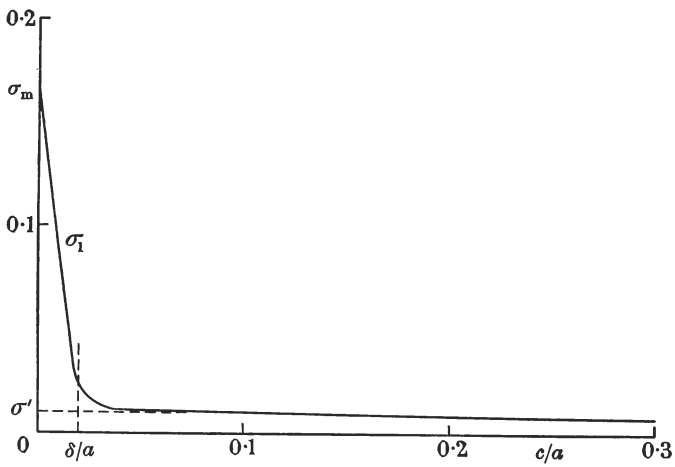

FIGURE 2. σ_1 as function of relative length c/a along σ_3 trajectory drawn from circle of contact. For definition of σ_m, σ' and δ see text.

To describe the progress of a crack along a path on which the prior stresses vary in such a manner as outlined above we note the distinctive behaviour of cracks under two extreme types of simple loading, namely (*a*) uniform tensile stress and (*b*) loading at the mouth only (as by the driving in of a wedge). Case (*a*) is the one treated by Griffith, and gives rise to an unstable critical crack length such that a shorter crack will not grow while a longer one will grow without limit. This critical crack length corresponds to a maximum in the sum of mechanical and surface energy. Case (*b*) results in a stable crack length corresponding to a minimum in the sum of these energies. In the light of this it is possible to foresee that the inhomogeneous stress distribution of the present case may give rise to four values of c at which the energy is stationary. First, a crack short compared with δ is in a condition similar to case (*a*)

and has a critical value c_0. As the load increases c_0 diminishes. If it becomes less than the size, c_f, of surface flaws already present, these should extend in depth, but not without limit because of the steep decrease in stress. The situation is now that of case (*b*), with a stable crack length c_1 of order of magnitude δ. At this stage we expect the crack to run around, deviating somewhat outside the σ_2 trajectory, to form a shallow ring-crack (*not* yet a cone crack). By hypothetical increase of c, which would not in these circumstances occur spontaneously, it would reach further into the region of lower, nearly constant, stress and increasingly resemble case (*a*) again, thus having a second unstable critical length c_2. If it could be made to exceed c_2 a further growth (at constant load) to a substantially greater depth should occur. The situation now, with the stress diminishing approximately as c^{-1}, and the crack extending on a widening front on the conical surface, is unlike either case (*a*) or (*b*) above, but we rely on Roesler's analysis (1956*b*) to confirm that it leads to a second stable length, c_3, substantially larger than a.

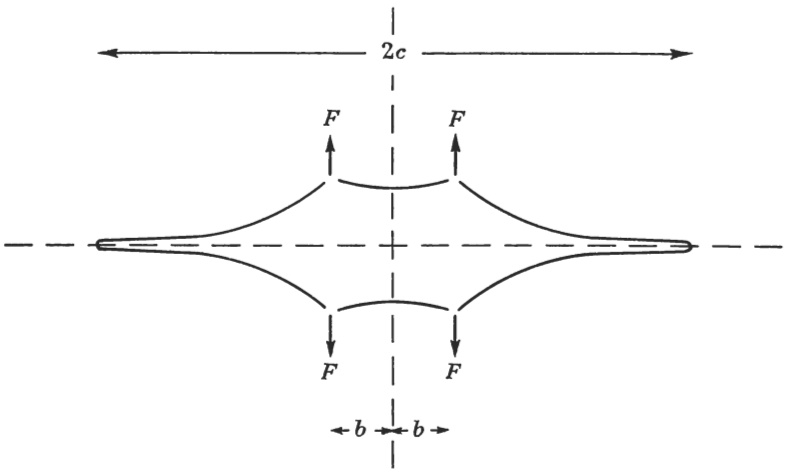

FIGURE 3. Crack, length 2*c*, in infinite medium, loaded with forces $\pm F$ at $\pm b$.

With a little further consideration we note that increasing the load on the indenter increases the stable crack lengths c_1 and c_3, and decreases the unstable crack lengths c_0 and c_2. Thus c_1 and c_2 mutually approach each other, until the corresponding maximum and minimum in the curve of total energy against c merge and vanish. At this stage the stable length c_1 becomes unstable and the crack spontaneously grows to the much greater new stable value c_3, forming the fully developed cone crack. Since, in this interpretation, the critical event for a fully developed fracture is the merging of c_1 and c_2, we will now pursue its significance in some quantitative detail.

The most severe approximation we need to make is to ignore both the curvature of the crack path and its widening into a cone, treating it as a plane crack normal to the surface. For c_0, c_1 and c_2 this approximation will be seen to be not too severe. To compute the condition that the crack will, at any stage, extend, we make recourse to some results developed by Irwin (1958).

The quantity $\mathrm{d}(\Delta U/|\mathrm{d}c|$ of § 3 above is represented in Irwin's notation by $\mathscr{G}$, designated the *strain energy release rate*. The principal standard cases for which

explicit calculations of $\mathscr{G}$ are available are cases in which the crack advances uniformly along its front, on a prescribed (in most cases, planar) path, so that the general infinitesimal increment of area, $\mathrm{d}c$, can be replaced by the increment, $\mathrm{d}c$, in a scalar quantity c representing the crack length, or half-length (for a double ended crack), or radius (for a crack of axial symmetry). $\mathscr{G}$ is then defined as $\mathrm{d}\Delta U/\mathrm{d}c$ *per unit length of crack front*: it is a function of the crack size, the applied load, and the size, shape, and elastic properties of the specimen. Irwin also defines a *stress intensity factor*, $\mathscr{K}$, of the dimensions stress $\times$ (length)$^{\frac{1}{2}}$, related to $\mathscr{G}$ (in the case of plane strain) by

$$\mathscr{G} = (\pi/E)(1-\nu^2)\mathscr{K}^2 \tag{4.3}$$

the virtue of which is that $\mathscr{K}$ values for a number of superposed loads are additive. To calculate $\mathscr{G}$ for the loading conditions specified by (4·1) we begin with a crack of length $2c$ (figure 3), infinitely long in the direction normal to the plane of the figure (so that a state of plane strain exists), embedded in an infinite, isotropic, elastic medium. The crack is subjected to localized forces $\pm F$ per unit length, acting normally to the crack plane, at positions $\pm b$. From Irwin (1958) we write the factor $\mathscr{K}_F$ for this loading

$$\mathscr{K}_F = \frac{2F}{\pi}\left(\frac{c}{c^2-b^2}\right)^{\frac{1}{2}}. \tag{4·4}$$

For any distribution of normal stress along the crack faces we may insert $F = \sigma(b)\,\mathrm{d}b$ into (4·4) and integrate. We now make use of the approximation introduced by Inglis (1913) that the rate of release of mechanical energy, as the crack extends, for a crack of length $2c$ in an infinite medium, is twice that for a surface crack of length c in a semi-infinite medium. The resulting value of $\mathscr{G}$ derived from (4·3) will therefore apply equally well to the latter case, which we are considering here. If we perform the integration of (4·4), inserting the values of stress given by (4·1), we obtain for $\mathscr{K}'$ and $\mathscr{K}''$ (corresponding to σ' and σ'' respectively):

$$\left.\begin{aligned}\mathscr{K}' &= \sigma' c^{\frac{1}{2}},\\ \mathscr{K}'' &= \frac{2}{\pi}(\sigma_{\mathrm{m}}-\sigma')c^{\frac{1}{2}}\left[\sin^{-1}\frac{\beta}{c}-\frac{c}{\delta}\{1-\sqrt{(1-\beta^2/c^2)}\},\right]\end{aligned}\right\} \tag{4·5}$$

where $\beta = c$ if $c < \delta$ and $\beta = \delta$ if $c \geqslant \delta$.

We are now in a position to evaluate the conditions for crack growth. We consider the following cases: (i) $c \ll \delta$, leading to an estimate of c_0; (ii) $\delta < c \ll a$, leading to c_1 and c_2; (iii) $a \ll c$, beyond which (4.1), and therefore (4·5), is no longer valid, and we resort to Roesler's estimate for c_3. The condition that the crack extends, according to the basic assumption of Griffith, is

$$\mathscr{G} \geqslant 2\gamma. \tag{4·6}$$

(i) *Evaluation of* c_0. Putting $c \ll \delta$ and integrating $\mathscr{K}'$ and $\mathscr{K}''$ over the length of the crack we reduce the sum $\mathscr{K} = \mathscr{K}' + \mathscr{K}''$ to

$$\mathscr{K} = \sigma_{\mathrm{m}} c^{\frac{1}{2}}\left[1-\frac{2c}{\pi\delta}\left(1-\frac{\sigma'}{\sigma_{\mathrm{m}}}\right)\right].$$

From (4·3) and (4·6) we have, ignoring terms in c^2/δ^2,

$$c^2-\frac{\pi\delta}{4(1-\sigma'/\sigma_m)}c+\frac{\gamma E\delta}{2\sigma_m^2(1-\nu^2)(1-\sigma'/\sigma_m)}=0,$$

from which

$$c=\frac{\pi\delta}{8(1-\sigma'/\sigma_m)}\left\{1\pm\left[1-\frac{32\gamma E(1-\sigma'/\sigma_m)}{\pi^2\sigma_m^2\delta(1-\nu^2)}\right]^{\frac{1}{2}}\right\}.$$

The solution corresponding to the positive sign alternative is excluded by the restriction $c \ll \delta$. Taking the negative sign and expanding yields

$$c_0=c_G\left\{1+\frac{4}{\pi}\left(1-\frac{\sigma'}{\sigma_m}\right)\frac{c_G}{\delta}+\ldots\right\}, \tag{4·7}$$

where

$$c_G=\frac{2E\gamma}{\pi\sigma_m^2(1-\nu^2)}$$

is the critical crack length according to Griffith for a uniform tensile stress σ_m. The correction terms become of importance when c_G/δ approaches unity, which occurs, for example, at light loads.

(ii) *Evaluation of c_1 and c_2.* Putting $c > \delta$ and integrating $\mathscr{K}'$ and $\mathscr{K}''$ over the length of the crack,

$$\mathscr{K}=\frac{(\sigma_m-\sigma')\delta}{\pi c^{\frac{1}{2}}}+\sigma' c^{\frac{1}{2}}.$$

Again, from (4·3) and (4·6), we arrive at a quadratic expression for c,

$$c^2-\frac{1}{\pi}\left[\frac{2\gamma E}{(1-\nu^2)\sigma'^2}-2(\sigma_m/\sigma'-1)\delta\right]c+\frac{1}{\pi^2}(\sigma_m/\sigma'-1)^2\delta^2=0.$$

Writing the coefficients of this quadratic as A_1 (= unity), A_2, and A_3 the solutions for c become

$$c_{2,1}=-\tfrac{1}{2}A_2[1\pm(1-4A_3/A_2^2)^{\frac{1}{2}}]. \tag{4·8}$$

It may be shown that the positive sign corresponds to an unstable crack length (c_2) and the negative sign to a stable length (c_1). The two roots of (4·8) merge (to become imaginary) when the square root term vanishes, that is, when $A_2^2=4A_3$. This condition reduces to

$$\sigma'^2\delta=\frac{\gamma E}{2(1-\nu^2)(\sigma_m/\sigma'-1)}. \tag{4·9}$$

Hence, from (2·1), (2·2), (4·1) and (4·2), writing P^* for the critical load at which $c_1=c_2=c^*$ (we shall use the asterisk to denote the value of any variable when $c_1=c_2=c^*$):

$$\left.\begin{aligned}\frac{P^*}{r}&=\frac{8\pi^2 a\sigma_m}{3\delta\sigma'(1-\sigma'/\sigma_m)}\frac{k\gamma}{(1-\nu^2)(1-2\nu)^2}\\&\simeq 2{\cdot}35\times10^5 k\gamma,\end{aligned}\right\} \tag{4.10}$$

in which the numerical factor is, of course, somewhat unreliable because of approximations in the treatment. It follows immediately from (4·8) and (4·9) that

$$c^*=-\tfrac{1}{2}A_2=\left(\frac{\sigma_m}{\sigma'}-1\right)\frac{\delta^*}{\pi}\simeq 5{\cdot}05\delta^* \tag{4·11}$$

which lies within the range $\delta < c \ll 50\delta$, as required.

We may test (4·10) by using it to calculate γ for the Pilkington plate glass studied by Tillett (1956) for which she found $P_c/r \simeq 6{\cdot}1 \times 10^7$ dyn/cm. Identifying P_c as P^*, with $k = \frac{2}{3}$ (steel indenter and glass specimen), (4·10) gives $\gamma = 4{\cdot}0 \times 10^2$ erg/cm^2. (Griffith found $\gamma \simeq 650$ erg/cm^2 and Roesler (1956*b*) 4000 erg/cm^2.)

According to our interpretation the shallow ring crack to depth c_1 passes unrecognized as fracture. This is, perhaps, not surprising since, with a steel indenter of $r = 1$ cm on a glass specimen, c_1 is calculated to be about 50 μm just prior to the merging of c_1 and c_2, and most experiments are performed with much smaller indenters. It is the merging of c_1 and c_2, allowing the crack to make the much greater extension to c_3, which we take to be the critical event producing the visible cone crack. This is consonant with Roesler's (1956*a*) proposed derivation of Auerbach's law by a size-energy scaling argument, but instead of applying it to the fully developed crack we would apply it to the crack at the critical merging of c_1 and c_2. The value of c^* at which this occurs is indeed proportional to a (equations (4·11) and (4·2)) and the approach to this critical condition is indeed reversible, whereas the evident weakness of Roesler's treatment was the seemingly necessary assumption of reversibility in the formation of the fully developed cone crack.

(iii) *Evaluation of* c_3. Equation (4·1) no longer satisfactorily describes the stress field about the crack path when $c \gg a$. Roesler finds, for loads greater than P_c, that the ratio $c_3/P^{\frac{2}{3}}$ tends to constancy. He evaluates this constant, which, when $\alpha = 68{\cdot}5°$ (see § 3), is given in our notation by

$$c_3 = 1{\cdot}6 \times 10^{-6} P^{\frac{2}{3}}, \tag{4·12}$$

when c_3 is expressed in centimetres and P in dynes.

5. Limits of validity of fracture criteria

By establishing P^* as a unit of load, and δ^* as a unit of crack length where, from (2·1), (4·2), (4·10),

$$\delta^* = 1{\cdot}17(k^2\gamma/E)^{\frac{1}{3}} r^{\frac{2}{3}} \tag{5·1}$$

for $\nu = \frac{1}{3}$, and by writing $\delta = \delta^*(P/P^*)^{\frac{1}{3}}$, we may conveniently reduce equations (4·7), (4·8) and (4·12) for crack length to the following expressions in P:

$$\left.\begin{aligned} \frac{c_0}{\delta^*} &= 4{\cdot}81 \times 10^{-2} \left(\frac{P^*}{P}\right)^{\frac{2}{3}} \left[1 + 8{\cdot}64 \times 10^{-2} \frac{P^*}{P} + \ldots\right] && (a), \\ \frac{c_{2,1}}{\delta^*} &= 9{\cdot}94 \left(\frac{P}{P^*}\right)^{\frac{1}{3}} \left[\frac{P^*}{P} - \frac{1}{2} \pm \left(\frac{P^*}{P}\right)^{\frac{1}{2}} \left(\frac{P^*}{P} - 1\right)^{\frac{1}{2}}\right] && (b), \\ \frac{c_3}{\delta^*} &= 4{\cdot}60 \times 10^2 \left(\frac{P}{P^*}\right)^{\frac{2}{3}} && (c). \end{aligned}\right\} \tag{5·2}$$

The crack lengths c in (5·2) are shown plotted in figure 4 as a function of P. In all cases the plot becomes inaccurate at $P \ll P^*$ and the curve is smoothed in this region. Also included in figure 4 is the variation of δ and a with P.

The qualitative description of crack formation given in § 4 may now be discussed in detail. In particular we note the dependence of behaviour of the crack on c_t, the

length of microscopic flaws present in the specimen. Let us consider the various regions of the curve in figure 4 within which c_f may fall.

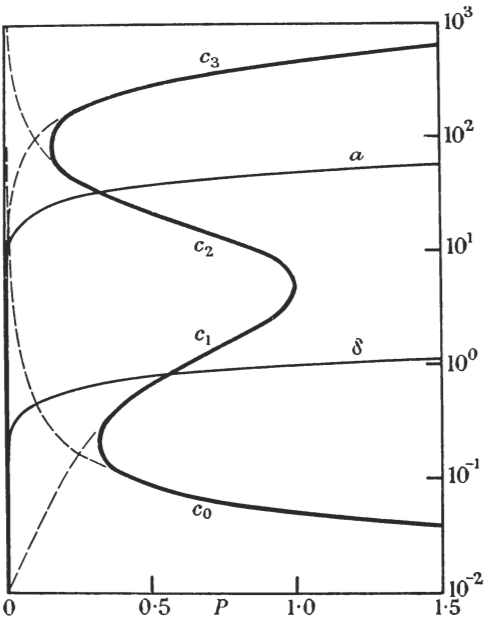

FIGURE 4. Plot of crack length c as a function of load P. δ^* and P^* are taken as units of length and load respectively. The broken lines at low P represent equations (5·2); the curve is smoothed in this region. Also shown is the variation of a and δ with P.

(i) $c_f < c_0^*$. The load P^* is in this case insufficient to make the crack unstable. A higher load $P_0 > P^*$ is required to reduce c_0 to length c_f, whereupon the crack grows directly to c_3, the length of the fully developed cone crack, in one stage. This behaviour will occur for indenters of radius larger than a limiting value r_u determined by the condition $c_0^* = c_f$. From (5·2a) and (5·1) this limiting condition becomes

$$c_f = c_0^* = 0{\cdot}052\,\delta^* = 0{\cdot}061(k^2\gamma/E)^{\frac{1}{3}} r_u^{\frac{2}{3}}, \tag{5·3}$$

that is,

$$r_u = 66(E/k^2\gamma)^{\frac{1}{2}} c_f^{\frac{3}{2}}. \tag{5·4}$$

r_u is thus an upper limit to the indenter radius for which Auerbach's law should be obeyed. Putting $k = \frac{2}{3}$, $E = 7 \times 10^{11}\,\mathrm{dyn/cm^2}$, $\gamma = 4{\cdot}0 \times 10^2\,\mathrm{erg/cm^2}$, (5·4) gives $r_u = 4{\cdot}0 \times 10^6 c_f^{\frac{3}{2}}$, with r_u and c_f in centimetres. Tillett found that Auerbach's law failed for indenters of larger radius than 3·5 cm. This would correspond to $c_f = 0{\cdot}9 \times 10^{-4}$ cm, which is of the right order of magnitude for Griffith flaws. According to the Griffith fracture criterion the tensile strength of the glass should then be $1{\cdot}4 \times 10^9\,\mathrm{dyne\,cm^{-2}}$. This is about seven times the momentary load strength 'at 1 % risk' for the glass Tillett used, quoted as 2500 to 3000 Lb (no doubt, per square inch): doubtless, this discrepancy must be attributed to the occasional much larger flaw.

Combining (5·2*a*) with (5·3), putting $c_0 = c_f$ and $P = P_c$, we obtain as a condition that a cone fracture develops

$$\frac{P^*}{P_c}\left(1+8{\cdot}64\times10^{-2}\frac{P^*}{P_c}+\ldots\right)^{\frac{3}{2}} = 1{\cdot}13\,\frac{r_u}{r},$$

or, approximately,

$$\frac{P_c}{P^*} = 0{\cdot}114+0{\cdot}886\,\frac{r}{r_u}. \tag{5·5}$$

Since P^* is itself proportional to r (equation (4·10)) this implies that to a first approximation, which becomes better as r increases, P_c is proportional to r^2, in accordance with a maximum stress criterion for fracture: but when r is not much larger than r_u the increase of P_c with r is somewhat slower. This is in excellent qualitative agreement with the results of Tillett, though the slope, $\mathrm{d}\log P_c/\mathrm{d}\log r$, of the best straight line through her points in the range $3{\cdot}5\,\mathrm{cm} < r < 11\,\mathrm{cm}$ is a little too low, about 1·8 rather than about 1·9 as we should expect. It is also in agreement with our interpretation that Tillett found that scratching the glass surface with a diamond (which we should regard as increasing c_f) had no effect on P_c for indenters of 2·5 cm radius or less but lowered P_c towards P^* for the larger indenters.

(ii) $c_0^* \leqslant c_f \leqslant c^*$. This is the range of validity of Auerbach's law. It extends over two powers of 10 in c_f (equations (4·11) and (5·3)) and accordingly over three powers of 10 in r. According as c_f is less than or greater than the value, about $\frac{1}{5}\delta^*$, corresponding to the first minimum of P (figure 4), the crack has or has not a small unstable growth to the corresponding stable depth c_1 when c_0 becomes equal to c_f. In either case, it then deepens stably with increasing load, still being a shallow ring crack, to (P^*, c^*) whereupon it extends unstably by a factor of order one hundred to the new stable value c_3 corresponding to a fully developed cone crack.

We have tended so far to ignore the fact that the area of contact increases with load. The first stable crack of depth c_1 therefore may not be the one which leads to the ultimate cone crack. It may be covered up, and thus be encompassed by the fully compressive region of the stress field before c_1 merges with c_2. Such a crack will close up and play no further part in the crack-formation process. However, as we expect this incomplete crack to be eccentric it should not be fully covered at once. So long as a substantial part of this crack remains uncovered it should reduce the magnitude of the surrounding tensile stress field (except near the edges where it intersects the circle of contact, where the tensile stress tends to reopen it against compression). This will tend to inhibit the formation of a second crack as a increases. If the first crack never attains the depth c^* a second will eventually initiate when the first becomes sufficiently well covered. There may therefore be one or more preliminary cracks within the final effective one. It is not quite clear what we should expect when a partially covered stable crack breaks through to form a fully developed cone fracture. The most strongly stressed portion is that just outside the circle of contact, and this doubtless makes the breakthrough first, then extending over the rest of the cone. If only a small part of the initial crack is covered when breakthrough occurs, the stress concentration at the edge of the remainder should be sufficient to reopen the covered portion. We should then see a ring crack actually

intersecting the circle of contact. If, on the other hand, a substantial part of the initial ring crack is covered when breakthrough occurs, one should expect the crack to extend from the uncovered portion only, then spread around the cone and grow back up to the surface. In this case the visible ring crack may lie entirely outside the circle of contact and may appear split on the side opposite to that at which breakthrough occurred.

(iii) $c_f > c^*$. From (4·11) and (5·1) this condition corresponds to

$$r < r_1 = 0{\cdot}72(E/k^2\gamma)^{\frac{1}{2}}c_f^{\frac{3}{2}} \tag{5·6}$$

so that r_1 represents a lower radius limit to the validity of Auerbach's law for a completely non-plastic material. Inserting the values of k, E and γ used in case (i) and $c_f = 0{\cdot}9 \times 10^{-4}$ cm into (5·6), we have $r_1 = 38 \times 10^{-4}$ cm. For smaller indenters, from equations (5·1), (5·2b) and (5·6), with $P = P_c$ and $c_f = c_2$, we have

$$\frac{r_1}{r} = 2{\cdot}83\left(\frac{P_c}{P^*}\right)^{\frac{1}{2}}\left[\frac{P^*}{P_c} - \frac{1}{2} + \left(\frac{P^*}{P_c}\right)^{\frac{1}{2}}\left(\frac{P^*}{P_c} - 1\right)^{\frac{1}{2}}\right]^{\frac{3}{2}} \tag{5·7}$$

so that, when $r < r_1$, $P_c (< P^*)$ tends toward proportionality with r^2 as in case (i).

One must not place too much reliance on these results since when c_f approaches the order of magnitude of a (cf. figure 4) the validity of the theory is distinctly doubtful. Moreover, as pointed out by Roesler (1956a) the Hertz stress formulae are untrustworthy at small r, when, under the critical load, a is no longer small compared with r, and furthermore plastic flow may occur even in a nominally brittle solid due to the very high stress produced under a small indenter (at P^* the maximum shear stress present according to the Hertz formulae is $12{\cdot}1(\gamma E^2/kr)^{\frac{1}{3}}$, which is approximately $10^{10}/r^{\frac{1}{3}}$ dyn/cm^2, r being in centimetres, with the parameters for glass with a steel indenter which we have used above). The case $c_f > c^*$ may arise more realistically for a specimen with relatively severe surface damage.

6. Discussion

In this section we take up a few corollary points. We were led into this subject by experimental observations on ring cracks in diamond (in which crystallographic cleavage introduces complications which we do not deal with here) and we have not performed systematic experimental observations on Hertzian fracture of glass. However, after reaching the conclusions of § 3 regarding the crack path we performed some impact experiments, dropping a relatively large steel sphere, of radius 0·875 cm, on to 1 in. thick plate glass from a height of *ca.* 17 cm, which gave approximately a 50 % probability of producing a cone fracture, the ball being greased so as to make the area of contact visible. Figure 5 is selected as a relatively simple representative result. In many cases there was more complex fracturing, but the following characteristics, seen without complication in figure 5, plate 2, are quite general: the crack trace is on one side quite close to the circle of contact (sometimes, as in this case, just inside it, but not always so), and striations visible on the subsurface crack have directions which appear to indicate that it has grown around in

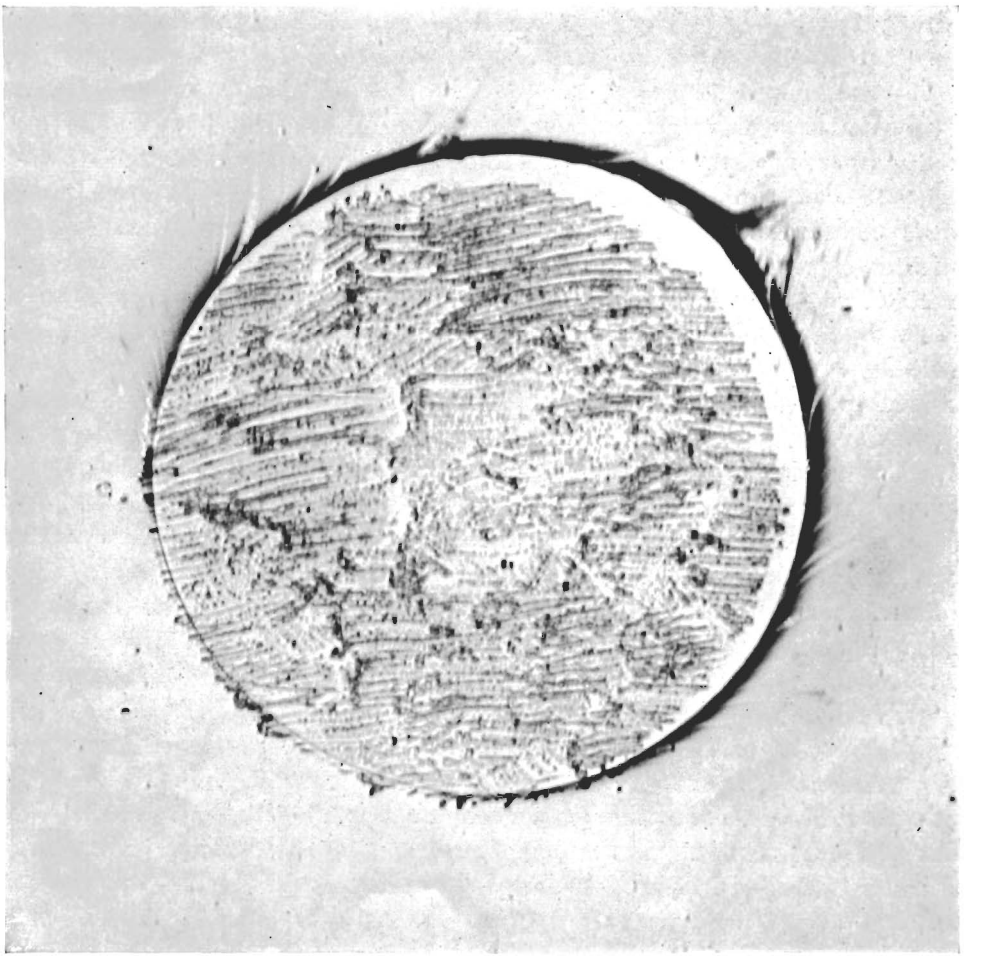

FIGURE 5. Surface view (reflected light) of a ring crack formed by dropping a steel ball, radius $r = 0{\cdot}875$ cm, onto a glass plate. The grease patch, radius $a = 0{\cdot}061$ cm, reveals area of contact.

(*Facing p.* 304)

both directions from this place of greatest proximity to the centre of the contact area, to meet itself at the opposite point of the ring. It would require more elaborate experiments than we have conducted to confirm the obliquity of the cone axis which we anticipate, and think we observe.

Our interpretation requires that at loads less than P_c, r being within the range for which Auerbach's law is valid, and especially when r is small compared with r_u, shallow ring cracks of depth c_1 should be produced. Since their penetration is less, by a factor of a hundred or more, than that, c_3^*, of the cone crack at critical load, we find no difficulty in supposing that they ordinarily pass unnoticed. However, apart from the prospect of detecting them with more delicate observation, we think they may have other observable consequences. In particular, repeated application of the same load would give opportunities for cumulative change in this crack, by entry of 'debris' (cf. Lawn & Komatsu 1966), or chemical change altering the effective value of γ, either of these causing gradual progress of the crack. This may provide an explanation of the observations of Hancox (1960) and Cooper (1961), quoted by Bowden & Tabor (1965), that with impact repeated a few thousand times a ring crack is produced in diamond at a lower load than is required in a single impact. Cooper looked for cumulative crack growth, and found instead that the crack appeared suddenly: but this is not incompatible with our expectation that it would extend suddenly by a factor of about 100 after growing slowly to the critical depth.

The reported factor 4, by which the maximum stress, σ_m, at fracture, is reduced (equivalent to a factor 64 in P for a fully elastic indenter) is too large for agreement with our calculations. However, we suspect that this factor may have been overestimated; it has been arrived at by combining the results for repeated impact tests made with steel balls with those for static tests made with a different (diamond) indenter, and no such large factor is deducible from the steel ball tests alone.

When the cone crack forms, the area within the ring at which it meets the surface is elastically depressed. If the ring crack is not much larger than the circle of contact, and especially if the applied load, or kinetic energy of impact, is more than the minimum necessary to make a cone crack, the indenter will then bear on the surface outside the ring. This should produce a secondary ring crack outside the primary one, but this time not spreading out in a conical skirt, because the former crack prevents the development of tensile stresses in the direction normal to it. The secondary crack should therefore join, or nearly join, the primary crack, to make a detachable collar of material around the neck of the cone.

Finally, we note that the Auerbach constant P^*/r, by equation (4·10), is proportional to γ and depends on elastic constants in a relatively insensitive way provided that the indenter is not elastically soft compared with the specimen (cf. equation (2·3)). Furthermore, it does not depend on the size of pre-existing cracks, or on anything else. It therefore affords, in principle, an excellent means of measuring the surface energy γ. This quantity, of course, may involve plastic work in the neighbourhood of the growing crack, and may be subject to chemical change. In so far as it is time-dependent, this method of measurement is quite appropriate, since the approach to the critical condition is in principle reversible, and the loading rate can be widely varied in a controlled manner. The disadvantages are that the numerical

constant in equation (4·10) has only been calculated on the basis of approximations, and must be subject to statistical variation if, as the evidence indicates, the precise crack path is subject to some statistical variation. Further changes in the numerical factor must occur in crystalline materials when crystallographic cleavage modifies the crack path. These disadvantages may be to some extent overcome, either by experimental 'calibration' of the method, or by more detailed computation. Even without this, the Auerbach constant appears to be a most valuable experimental quantity bearing on the strength of materials subject to brittle fracture, primarily related to γ, and the extension of the Hertzian fracture test to a wider range of materials may be advocated.

Note added 10 October 1966

A referee has advised us to give reference to the review paper by G. I. Barenblatt (1962, *Adv. Appl. Mech*, **7**, 56) in which cracks with ranges of stable and unstable growth are discussed, with other specific examples than that discussed here, and to the paper of D. M. Marsh (1964, *Proc. Roy. Soc.* A **279**, 420) whose evidence of plastic flow in glass is of importance in connexion with the failure of Auerbach's law at the lower limit of indenter size.

One of us (B. L.) wishes to thank Industrial Distributors Limited for a Research Fellowship during the tenure of which this work was done.

References

Auerbach, F. 1891 *Ann. Phys. Chem.* **43**, 61.
Bowden, F. P. & Tabor, D. 1965 *Physical properties of diamond*. Edited by R. Berman. Oxford: Clarendon Press.
Cooper, R. 1961 Thesis, Cambridge University.
Fuchs, S. 1913 *Phys. Z.* **14**, 1282.
Griffith, A. A. 1920 *Phil. Trans.* A **221**, 163.
Griffith, A. A. 1924 *First Internat. Congr. App. Mech.* (*Delft*), 55.
Hancox, N. L. 1960 Thesis, Cambridge University.
Haward, R. N. 1949 *The strength of plastics and glass*. London: Cleaver-Hume.
Hertz, H. 1881 *J. reine angew. Math.* **92**, 156. Reprinted, in English, in *Hertz's Miscellaneous papers*, ch. 5. London: Macmillan.
Hertz, H. 1882 *Verhandlungen des Vereins zur Beförderung des Gewerbe fleisses*, **61**, 449. Reprinted, in English, in *Hertz's Miscellaneous papers*, ch. 6. London: Macmillan.
Inglis, C. E. 1913 *Trans. Inst. Naval Archit.* **55**, 219.
Irwin, G. R. 1958 *Handb. Phys.* **6**, 551. Berlin: Springer.
Lawn, B. R. & Komatsu, H. 1966 *Phil. Mag.* **14**, 689.
Love, A. E. H. 1927 *A treatise on the mathematical theory of elasticity*. Cambridge University Press.
Morton, W. B. & Close, L. J. 1922 *Phil. Mag.* **43**, 320.
Roesler, F. C. 1956*a* *Proc. Phys. Soc.* B **69**, 55.
Roesler, F. C. 1956*b* *Proc. Phys. Soc.* B **69**, 981.
Tillet, J. P. A. 1956 *Proc. Phys. Soc.* B **69**, 47.

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRINTING HOUSE, CAMBRIDGE